

Using action dynamics to assess competing stimulus control during stimulus equivalence testing

Denis P. O’Hora · Ian T. Tyndall · Mairéad McMorro · Rick A. C. Dale

Published online: 3 February 2013
© Psychonomic Society, Inc. 2013

Abstract Previous studies have identified potential sources of competing stimulus control in tests for stimulus equivalence. The present experiment employed the Nintendo Wii remote (Wiimote®) to investigate whether such competition would affect suboperant action dynamics (e.g., topographies of equivalence responses). Following one-to-many training on conditional discriminations sufficient to establish three 3-member equivalence classes, participants were presented with a test for equivalence responding that included five different trial types. These included “traditional” equivalence trials, on which the incorrect stimulus had previously been presented as a correct comparison stimulus during training, trials on which a novel unrelated word was provided as the incorrect comparison, and trials on which the incorrect stimulus varied in orthographical and phonological similarity to the sample stimulus. The presence of phonological and orthographic distractor stimuli significantly reduced the probability of equivalence-consistent responding, relative to neutral distractors, but this effect was almost exclusively seen in participants who failed to demonstrate equivalence on traditional equivalence trials. Analyses of correct response trajectories suggested that the prior history of reinforcement for choosing the incorrect stimulus on the traditional equivalence trial gave rise to greater competition than did phonological or orthographic similarity between the sample and incorrect comparisons.

Keywords Stimulus control · Phonology · Orthography · Stimulus equivalence · Action dynamics · Humans · Stimulus control topography coherence · Naming theory · Relational frame theory

Stimulus equivalence constitutes a behavioral phenomenon that may, in part, account for the development of symbolic relations among stimuli (e.g., Barnes, 1994; Dougher, 1998; Sidman, 1994, 2000). Stimulus equivalence is typically examined in a matching-to-sample (MTS) procedure using conditional discriminations. For example, in the presence of an arbitrary stimulus A1 (sample), choosing another stimulus B1 (comparison) is reinforced. Then, in the presence of B1 (sample), choosing C1 (comparison) is reinforced. Stimulus equivalence requires that, in the presence of A1, participants choose C1 (transitivity) and can reverse both of the trained relations (choose A1 in the presence of B1 and C1 in the presence of B1; symmetry). Choosing A1 as a comparison when C1 is a sample (combined symmetry and transitivity) is often employed as the test of the emergence of the equivalence class (e.g., Sidman & Tailby, 1982).

There are a number of competing explanations of stimulus equivalence and the role it plays in accounting for complex behavior (e.g., Dube & McIlvane, 2003; Hayes, Barnes-Holmes, & Roche, 2001; Horne & Lowe, 1996; Sidman, 1994, 2000; Tonneau, 2001; Tonneau & Gonzalez, 2004). Studies testing the relative efficacy of such differing accounts in accounting for new empirical data (e.g., Minster, Elliffe, & Muthukumaraswamy, 2011; Minster, Jones, Elliffe, & Muthukumaraswamy, 2006) highlight that there is still much to learn in this area and that research employing the stimulus equivalence paradigm can still play a critical role in helping to isolate and identify key processes that underlie complex human behavior. One such area that requires further exploration is the issue of stimulus control exerted by sample and comparison stimuli in typical MTS

D. P. O’Hora (✉) · M. McMorro
School of Psychology, National University of Ireland Galway,
University Road,
Galway, Ireland
e-mail: denis.ohora@nuigalway.ie

I. T. Tyndall
University of Chichester, Chichester, UK

R. A. C. Dale
University of California Merced, Merced, CA, USA

procedures. Stimulus control topography coherence theory (SCTC; Dube & McIlvane, 2003) is particularly concerned with such issues. SCTC proposes that equivalence-consistent responding requires coherence between the stimulus properties and behavioral history (stimulus control topographies) that are observed to control the participants' responses and experimenter-defined stimulus control topographies. The present study sought to identify competing sources of stimulus control that are of particular relevance when written words are employed as stimuli.

Previous research has shown that the presence of incorrect comparison stimuli that are formally or functionally similar to the sample stimulus presented in equivalence testing reduces the probability of demonstrating the expected equivalence responses. For example, Stewart, Barnes-Holmes, Roche, and Smeets (2002) trained and tested for the emergence of three 3-member equivalence classes comprising nonsense syllables as stimulus elements that varied in color across conditions. Stewart et al. demonstrated that participants were less likely to show equivalence-consistent responding when incorrect comparison stimuli were presented in the same color as the sample. Thus, formal or topographical similarities between stimuli interfered with equivalence-consistent responding.

When sample and incorrect comparison stimuli have a similar function, the relationship between sample and positive comparison may be disrupted despite class-consistent baseline relations. Tyndall, Roche, and James (2004) employed a discrimination-training procedure to establish S+ functions for six stimuli (nonsense syllables) and S− functions for a further six stimuli. In general, participants required more test trial blocks to demonstrate the emergence of stimulus equivalence classes when the sample and correct comparison stimuli were discriminative for different operant responses and the sample and incorrect comparison stimuli were discriminative for the same basic operant response. Such studies serve to highlight the importance of identifying potential competing sources of stimulus control during equivalence testing.

Reductions in equivalence-consistent responding attributed to competing stimulus control in equivalence testing have also been observed in studies that employed real words as stimuli (e.g., Leslie et al., 1993; Moxon, Keenan, & Hine, 1993; Plaud, 1995, 1997; Watt, Keenan, Barnes, & Cairnes, 1991). For example, Watt et al. reported that a sample of participants from Northern Ireland, where communities are largely segregated along religious lines, found it difficult to demonstrate equivalence-consistent responding when Catholic names and Protestant symbols were presented as sample and correct comparison stimuli. However, English participants, who were unfamiliar with these cultural markers, had little difficulty forming the appropriate equivalence classes. Moxon et al. also reported marked impairments in

equivalence-consistent responding when male names and female-stereotypic job roles were provided as sample and correct comparison stimuli and vice versa. Similar reductions in equivalence-consistent responding by persons diagnosed with generalized anxiety disorder were observed when pleasant-state adjectives (e.g., *relaxed*) and anxiety-provoking terms (e.g., *public speaking*) were presented as sample and correct comparison stimuli (Leslie et al., 1993). The probability of equivalence-consistent responding is thus reduced when *correct* comparison stimuli have *opposing* stimulus functions to the sample stimulus and when *incorrect* comparison stimuli have *similar* functions or “meaning” as the sample stimuli. That is, preestablished functions may lead to responding in accordance with functional equivalence, rather than in accordance with the predicted equivalence relations.

The interference effects observed in the foregoing studies suggest that verbal behavior may play a role in facilitating or obstructing the formation of stimulus equivalence classes. Corroborating evidence for this position has been reported in a number of further studies (e.g., Arntzen, 2004; Dugdale & Lowe, 1990; Horne & Lowe, 1996, 1997; Lowe, Horne, Harris, & Randle, 2002; Mandell & Sheen, 1994; Miguel, Petursdottir, Carr, & Michael, 2008; Randell & Remington, 1999, 2006). Arntzen varied the nameability and familiarity levels of stimuli presented to adult participants and found that increased levels significantly increased rates of equivalence class formation. Employing children as participants, Dugdale and Lowe demonstrated that teaching a class-consistent common name for a set of visual stimuli facilitated subsequent equivalence-consistent responding in children who had previously failed to demonstrate the expected equivalence classes comprising the visual stimuli. These findings are in line with Horne and Lowe's (1996, 1997) proposition that a particular verbal behavior, naming, may provide both the necessary and sufficient preconditions for equivalence class formation (cf. Luciano, Begera, & Valverde, 2007).

Of particular relevance to the present research, Randell and Remington (1999) presented adult participants with different class arrangements of the same easily nameable but formally unrelated pictorial stimuli. The pictorial stimuli presented included no perceptual or topographical similarities that provided any consistent basis for categorization. However, for one group of participants, the names of the stimuli composing the classes rhymed with each other. Randell and Remington (1999) reported that despite the absence of instructions to name stimuli, participants named the stimuli presented (p. 407). Moreover, the probability of equivalence-consistent responding was enhanced when the stimuli in a class rhymed with each other (e.g., *boat* and *goat*) relative to when they did not rhyme (e.g., *boat* and *chair*). Subsequently, in generalized class

formation trials, all participants previously exposed to classes of rhyming stimuli selected only rhyming comparisons, while almost half of the participants previously exposed to classes composed of nonrhyming stimuli also consistently selected rhyming comparisons. Randell and Remington (2006) replicated and extended the previous study, stating that “when a ready verbal basis for stimulus categorization is available, that basis can become functional, even in the absence of specific training” (p. 338).

The interaction between orthographic and phonological information during written word processing was investigated in experiments conducted by Polich, McCarthy, Wang, and Donchin (1983) that involved a series of matching tasks in which word pairs rhymed and looked alike, rhymed but looked different, looked similar but did not rhyme, or did not rhyme and did not look alike. Polich et al. found that both orthographic and phonological stimulus properties seemed to multiplicatively affect error rates and response speed measures in tasks that solely required an orthographic or a phonological response. These researchers proposed that their data supported a dual-route model of reading, a lexical route and a nonlexical route (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). That is, one route from word to meaning is dependent on private phonological responses, and one is not. Within behavior analysis, relational frame theory (RFT; Hayes et al., 2001) also proposes that phonological responses may or may not enhance equivalence-consistent responding depending on the particular behavioral histories that support such responding (e.g., Luciano et al., 2007).

The potential impact of private phonological responses (e.g., textual responses; Skinner, 1957) on equivalence-consistent responding has implications for understanding some of the processes involved in reading. Cognitive and developmental theories of reading generally propose that orthographic, phonological, and semantic representations are activated during reading (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart et al., 2001; Plaut, McClelland, Seidenberg, & Patterson, 1996), and many behavior analysts (e.g., Hayes et al., 2001; Horne & Lowe, 1996; Sidman & Tailby, 1982) agree that reading requires equivalence-consistent responding. If so, reading written words seems particularly susceptible to the type of interference effects mentioned previously, because different words often share topographical (orthographical) features and functions (phonological or semantic). Written words with very different meanings may be topographically similar in that they share letters and sequences of letters (e.g., *clown*, *thrown*), and they may be functionally similar in that their presence occasions similar sounds (e.g., homophones such as *madel-maid* and *chased/chaste*). In the case of homonyms, the same word (lexical entry) may even differ in meaning depending on context (e.g., *sit by the fire*, *fire a gun*). The present

study sought to ascertain whether such sources of stimulus similarity interfere with the production of equivalence-consistent responding.

The effects of competing stimulus control are typically observed in lower response accuracy (i.e., probability of the expected response) or increased response latencies. Recent research in cognitive science has suggested that action dynamics provide further indices of such competition. Action dynamics refer to changes in the topography of the operant response over time. For example, it is possible to measure the trajectory of the operant response. Spivey, Grosjean, and Knoblich (2005) conducted a study in which participants were required to choose a named object from a set of two stimuli. Participants clicked a stimulus at the bottom of a computer screen, using a computer mouse to hear a name (e.g., *Candy*; an auditory sample stimulus), then moved the cursor up the screen to choose between two visual stimuli (e.g., a wrapped candy or a wax candle; visual comparison stimuli). On trials on which the name of the incorrect visual stimulus (e.g., the wax candle) was phonologically similar to the auditory stimulus (e.g., *Candy*), response trajectories showed greater curvature toward the incorrect stimulus prior to choosing the correct stimulus. The curvature of the response trajectories thus provided an index of competition between the two response alternatives. That is, the incorrect comparison exerted greater stimulus control when the verbal name associated with the visual stimulus shared characteristics with the auditory sample stimulus. In addition, this response trajectory effect demonstrated that competition between responses persisted *after* the response had been initiated.

Further evidence that response trajectories may highlight competing stimulus control was provided by Dale, Roche, Synder, and McCall (2008), who conducted two experiments to investigate action dynamics during MTS conditional discrimination learning. Participants were presented with unfamiliar symbols; one shape was designated as the pair’s sample, the other the pair’s match/comparison stimulus. Dale et al. employed the Nintendo Wiimote© to track the trajectory of participants’ behavioral response patterns (i.e., arm and hand movements) while they selected one of two comparison stimuli in the MTS paired associates task. Participants moved the Wiimote more quickly and more steadily and also began to press more firmly to choose stimuli as they became familiar with the MTS task and the stimuli that predicted reinforcement. These action dynamics grew more fluent and consistent over the duration of the task, and these changes correlated with the probability of choosing the correct response. These findings demonstrated that body movements (response topographies) changed systematically with overall response probability (i.e., learning of stimulus pairings).

The present research continues a long tradition of behavior-analytic research that has linked stimulus equivalence and

derived relational responding with verbal behavior, particularly in the case of reading (e.g., Sidman, 1971; Sidman & Cresson, 1973). The present study examined potential sources of conflicting stimulus control during stimulus equivalence testing. Following conditional discrimination training, participants were exposed to five types of trials during MTS equivalence testing to examine the effects of sample–comparison orthographic and phonological similarity. The incorrect comparison stimulus was (1) a stimulus that had previously been presented as a correct comparison stimulus on some MTS trials, (2) a neutral novel stimulus that was orthographically and phonologically distinct from the sample, (3) a novel stimulus that was orthographically similar to but phonologically distinct from the sample, (4) a novel stimulus that was phonologically similar to but orthographically distinct from the sample, or (5) a novel stimulus that was both orthographically *and* phonologically similar to the sample stimulus. It was expected that equivalence-consistent responding would be less likely to occur on test trials that included orthographically and/or phonologically similar incorrect comparisons than on control trials, in which incorrect comparisons were neither phonologically nor orthographically similar. In addition, it was expected that the presence of orthographically and/or phonologically similar incorrect comparisons would increase correct response times and increase the curvature and complexity of correct response trajectories.

Method

Participants

Twenty-four (17 females, 7 males; mean age, 20 years) university undergraduates of psychology at the National University of Ireland, Galway participated in this experiment. The participants were unfamiliar with the phenomenon of stimulus equivalence but were fully debriefed at the end of the experiment. English was the first language of all participants, and none was diagnosed with language deficits. Students received course credit for participation. All participants were right-handed.

Interface display and device

Stimuli were presented via a data projector. Participants stood behind a small table in a square marked on the ground. The projector was firmly attached to the ground below the table, pointing toward the wall (Fig. 1). Experimental tasks were coded in PsyscopeX B51 (Bonatti, n.d.; Cohen, MacWhinney, Flatt, & Provost, 1993) running on an Apple Mac Mini©. The Apple Mac Mini's display was projected onto a flat white wall, approximately 3.5 m from where the participant stood. The Nintendo Wiimote© was used as a

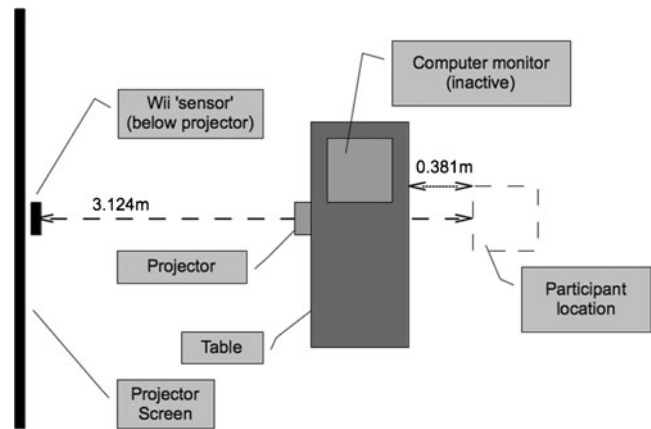


Fig. 1 Schematic top-down sketch of the experimental room, including the positioning of apparatus and the distance (in meters) from the participant to the projector screen. A data projector set beneath a table in front of the participant projected an image driven by the computer at the projector screen

wireless, arm-extended pointing device. The Wiimote was paired with the Mac Mini through the Bluetooth transfer protocol, and a Macintosh framework called DarwinRemote© (Kimura, n.d.) was employed to translate Wiimote movement into on-screen cursor movement. Under the projection screen, an infrared LED (“sensor bar”) was placed to calibrate the Wiimote. Participants stood behind the small table with the Wiimote in the dominant hand and with their arm extended comfortably. Lights were dimmed during the experiment. All responses were made using the Wiimote and PsyscopeX recorded responses and cursor position (x -, y -coordinates) at approximately 50 Hz (every 20 ms).

The conditional discrimination training and testing phases (phases 1–4, described below) employed auditory stimuli (spoken words) as stimuli. These stimuli were delivered to the participants through headphones.

Materials

Twenty-one stimuli were employed in Experiment 1 (see Table 1). Three stimuli were nonsense syllables (“CUG,” “ZID,” and “JOM”). The remaining 18 stimuli were English one-syllable words taken from the MRC Psycholinguistic database (Wilson, 1988), which were standardized for familiarity ($M = 520$, $SD = 22$). The stimuli were assigned alphanumeric labels for ease of communication, although the participants did not see these designations. The three nonsense syllables were presented as auditory stimuli during training, spoken by an Irish male experimenter and recorded in Audacity (Audacity Development Team, n.d.) for presentation by the experimental program. All visual stimuli were presented on a 15-in. screen in black Comic Sans MS 18-point font on a white background. Feedback, in the form of “CORRECT” or “WRONG,” was presented in the center of

Table 1 The 21 stimuli employed across phases 1–5 in Experiment 1 (with word familiarity figures for the English one-syllable words in parentheses)

| | A | B | C | Distractor comparisons |
|---|-----|-------------|------------|--|
| 1 | Cug | Grow (548) | Art (529) | D1po: Flow (518) D1p: Toe (578) D1o: Cow (529) D1n: Base (520) |
| 2 | Zid | Rear (547) | Trip (523) | D1po: Spear (513) D1p: Deer (509) D1o: Bear (526) D1n: Dawn (507) |
| 3 | Jom | Stone (564) | Nest (521) | D1po: Cone (506) D1p: Moan (510) D1o: None (569) D1n: Pick (524) |

the screen in capital letters in Times New Roman 26-point font.

Procedure

Training and testing for stimulus equivalence relations employed a one-to-many MTS training design (Arntzen & Holth, 1997) consisting of five phases, outlined in Fig. 2.

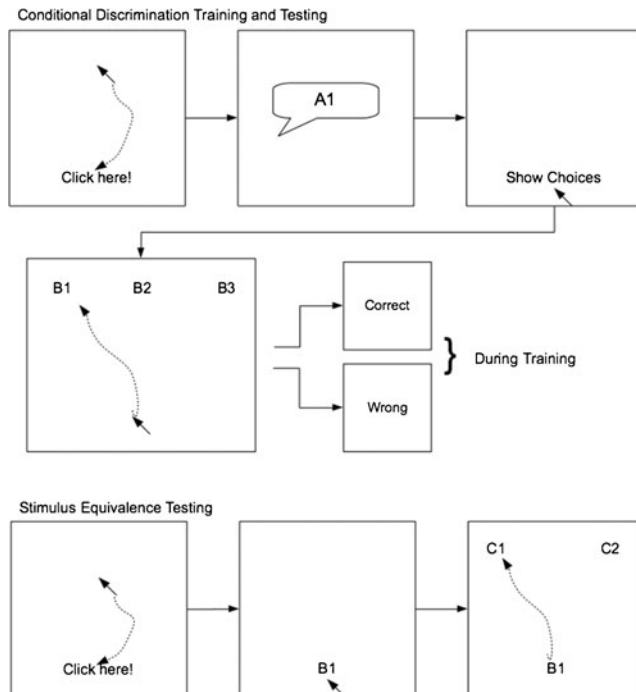


Fig. 2 Diagrammatic representations of the procedure during conditional discrimination training and testing and during the test for stimulus equivalence. During conditional discrimination training, auditory stimuli were presented as sample stimuli, but stimulus equivalence testing required visual–visual conditional discriminations

Phases 1–4 provided one-to-many training necessary to produce three 3-member equivalence classes (i.e., A1–B1–C1, A2–B2–C2, and A3–B3–C3). The A stimuli constituted the nodal stimuli and were presented as auditory stimuli. These phases consisted of blocks of MTS training and test trials. Following these four phases, participants were exposed to a stimulus equivalence test, in which stimulus equivalence probes were mixed with distractor probes.

Phases 1 and 2: Training and testing A–B conditional discriminations

Participants were read the following instructions prior to starting the A–B training and testing phases.

Welcome to this study and thank you for your participation.

Please put on the headphones and stand inside the box, which is marked on the floor.

To begin, please click on “Click Here” using the A button on the Wii remote. You will now hear a spoken word, when you have heard this word click on “Show Choices” using the A button on the Wii remote. Now choose one of the three words which appear on the screen, again using the A button on the Wii remote. For the first few trials, you will receive feedback in the form of “CORRECT” or “WRONG”—this is the training phase. Once you have successfully completed this training phase, you will immediately move on to the testing phase, this is the very same as the training phase but without feedback.

The experimenter checked whether the participant could hear the auditory stimuli and adjusted the volume of the computer if necessary. Participants were asked to inform the experimenter if, at any time, they could not hear the auditory stimuli. Phase 1 consisted of three conditional discriminations (A1–B1/B2/B3, A2–B1/B2/B3, and A3–B1/B2/B3; italics denote the correct comparison) presented 6 times each in quasirandom order within a block of 18 trials. At the beginning of each trial, the words “click here!” appeared in the bottom center of the projection area. Once participants clicked this stimulus, the auditory stimulus was played in the participants’ headphones.

Following an interstimulus-interval of 1 s, three comparison stimuli (e.g., B1, B2, and B3) were presented as comparisons at the top of the projection area. Participants chose one of three comparisons using the Wiimote by clicking the response button. Correct responses (e.g., choosing B1 in the presence of A1) were followed by the presentation of the word “CORRECT” in green in the center of the screen for 1 s, and incorrect responses were followed by “WRONG” in red for the same duration. An intertrial interval (ITI) of 1 s elapsed before the beginning of the next trial. Trials were

presented in a quasirandom order in blocks of 18 trials, which controlled for number of presentations of each sample and location of comparisons. Incorrect comparisons on a particular trial were subsequently employed as correct comparisons on other trials within the block.

The mastery criterion required that participants choose the correct comparison on at least 16 of the 18 trials in a block of training trials before progressing to phase 2. If participants failed to meet the mastery criterion, they were exposed to a further block of 18 trials. Participants were exposed to a maximum of 15 trial blocks in phase 1. In phase 2, participants were exposed to the same 18 trials in a quasirandom order without feedback. The mastery criterion remained at 16 out of 18 trials in a block. If the participants satisfied the criterion, they proceeded to phase 3. If not, the participants were reexposed to phase 1 A–B conditional discrimination training.

Phases 3 and 4: Training and testing A–C conditional discriminations

Following successful completion of phase 2, phase 3 was presented without a break (participants were, however, allowed to take a break at any time if they wished to). Phase 3 consisted of three further conditional discriminations (A1–C1/C2/C3, A2–C1/C2/C3, and A3–C1/C2/C3) presented 6 times each in a quasirandom order within a block of 18 training trials. Trial and feedback presentation followed the same procedure as phase 1. If a participant satisfied the mastery criterion of 16 correct out of the 18 trials in a block, they proceeded to phase 4 testing A–C conditional discriminations. Phase 4 presented the same 18 trials in the absence of feedback, and the mastery criterion remained at 16 out of 18 test trials correct. If participants satisfied the mastery criterion, they proceeded to phase 5 testing for stimulus equivalence. If not, the participants were reexposed to phase 3 training A–C conditional discriminations, subject to a maximum of 15 blocks.

Phase 5: Testing for stimulus equivalence (B–C)

Prior to beginning phase 5, the following instructions were presented.

All words in this phase will be presented visually onscreen.

In the same way as Phase 1 and 2, click on “Click Here” to begin. A word will then appear on screen. When you have read the word, click on it. Two words will then appear on screen. You must choose one of these by clicking on it using the A button on the Wii remote. There will be no training, so no feedback will be given at any stage during this phase.

Phase 5 tested for responding in accordance with derived combined symmetry and transitivity (equivalence) relations between B and C stimuli, using a two-comparison MTS

procedure under varying conditions of competing stimulus control (see Fig. 3). These trials employed a two-choice procedure in order to facilitate competition between the distractor and the correct comparison. B stimuli were presented as samples, and C stimuli were presented as comparisons. At the beginning of each trial, the sample B stimulus was presented at the bottom center of the projection area. On clicking the sample stimulus, two comparison stimuli were presented in the top left and right corners of the screen, and participants chose by moving the cursor to one of these comparisons and clicking a button on the Wii remote. No feedback was provided, and an ITI of 1 s preceded the subsequent test trial. Trials were presented in a quasirandom order in blocks of 36 (12 equivalence trials and 24 distractor trials). The experimental software controlled the number of presentations of each sample, correct comparison, and incorrect comparison stimulus and the locations of the comparison stimuli. Participants were not required to choose B stimuli as comparisons in the presence of C stimuli (C–B relation). First, B–C derived relations required both symmetry and transitivity, because we employed a one-to-many training design. Second, designing C–B test trials would have required a further 12 distractor stimuli (4 per class) controlled for frequency and semantic associations (see below).

Five types of test trials were employed: (1) stimulus equivalence test trials, on which the incorrect comparison was one of the other C stimuli presented during the conditional discrimination training and testing phases; (2) trials on which the incorrect comparison was neither phonologically nor orthographically similar (e.g., *grow* and *base*; neutral [N] trials); (3) trials on which the incorrect comparison was only orthographically similar (i.e., *grow* and *cow*; orthographic-only [O] trials); (4) trials on which the incorrect comparison was only phonologically similar (e.g., *grow* and *toe*; phonological-only [P] trials); or (5) trials on which the incorrect comparison stimulus was both phonologically and orthographically similar to the

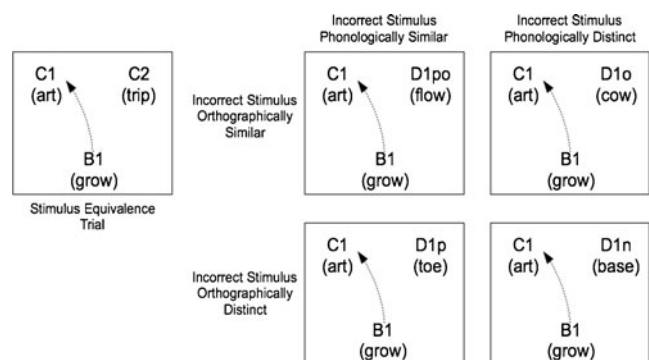


Fig. 3 Diagrammatic example of a stimulus equivalence trial type and the four distractor trial types examined in phase 5 equivalence testing: phonologically and orthographically similar comparison stimulus (PO); phonologically similar only (P); orthographically similar only (O); neither phonologically nor orthographically similar (N)

sample stimulus (e.g., *grow* and *flow*; PO trials). So, if B1 (*grow*) was presented as a sample, the correct comparison was C1 (*art*), and the four possible incorrect comparisons were designated D1p (*flow*), D1p (*toe*), D1o (*cow*), and D1n (*base*; see Fig. 3). For each B stimulus, there were two possible stimulus equivalence trials, one for each of the remaining C stimuli (e.g., choose C1 or C2; C1 or C3). These stimuli were presented on either corner, giving 4 trials per stimulus and 12 equivalence trials in total. The remaining four trial types (N, O, P, and PO) were presented twice each, generating a total of 8 test trials for each of the three B stimuli, resulting in a total of 24 distractor trials and a grand total of 36 trials in phase 5. Trials were presented in a quasirandom order such that all trials were presented but in no particular order.

Following phase 5, participants were fully debriefed as to the rationale of the experiment, thanked for their participation, and awarded credits.

Results

Overall performance of each participant across phases 1–5 is provided on the left-hand side of Table 2, and performances of each participant in each condition during stimulus equivalence testing are presented on the right. Of 24 participants, 17 (70.8 %) were deemed to have demonstrated the formation of stimulus equivalence classes (greater than 80 % correct responses on stimulus equivalence probes). The number of training trials required to satisfy the mastery criterion in A–B training ranged from a minimum of 18 trials (one block; 6 participants) to a maximum of 108 trials (six blocks; participant 17). With the exception of participant 17, all participants demonstrated the required A–B conditional discriminations within 72 trials (four blocks). Participants were exposed to just one block of A–C training trials prior to testing, due to limitation of the software, but

Table 2 Performance data of the 36 participants in Experiment 1 across phase 1 A–B training, phase 2 A–B test, phase 3 A–C training, phase 4 A–C test, and phase 5 B–C equivalence test

| | A–B | A–B Test | A–C | A–C Test | B–C Test | Accuracy across conditions | | | | |
|----|-----|----------|-----|------------|----------|----------------------------|-------|-------|-------|--------|
| | | | | | | SE (%) | N (%) | O (%) | P (%) | PO (%) |
| 1 | 54 | 18/18 | 18 | 18/18 | 14/36 | 60* | 83 | 0 | 33* | 0 |
| 2 | 72 | 18/18 | 18 | 18/18 | 36/36 | 100 | 100 | 100 | 100 | 100 |
| 3 | 72 | 17/18 | 18 | 18/18 | 20/36 | 18* | 100 | 67* | 50* | 67* |
| 4 | 36 | 18/18 | 18 | 18/18 | 35/36 | 100 | 100 | 100 | 83 | 100 |
| 5 | 18 | 18/18 | 18 | 18/18 | 9/36 | 58* | 33* | 0* | 0* | 0* |
| 6 | 72 | 18/18 | 18 | 17/18 | 34/36 | 83 | 100 | 100 | 100 | 100 |
| 7 | 54 | 18/18 | 18 | 18/18 | 36/36 | 100 | 100 | 100 | 100 | 100 |
| 8 | 36 | 18/18 | 18 | 18/18 | 35/36 | 100 | 100 | 100 | 100 | 83 |
| 9 | 18 | 16/18 | 18 | 17/18 | 35/36 | 100 | 100 | 80 | 100 | 100 |
| 10 | 18 | 17/18 | 18 | 18/18 | 13/36 | 75* | 33* | 33* | 0* | 0* |
| 11 | 54 | 18/18 | 18 | 18/18 | 30/36 | 82 | 100 | 100 | 60* | 100 |
| 12 | 36 | 17/18 | 18 | 18/18 | 11/36 | 55 | 67* | 0* | 0* | 0 |
| 13 | 36 | 18/18 | 18 | 17/18 | 35/36 | 100 | 100 | 100 | 100 | 80 |
| 14 | 18 | 18/18 | 18 | 18/18 | 36/36 | 100 | 100 | 100 | 100 | 100 |
| 15 | 36 | 18/18 | 18 | 18/18 | 35/36 | 100 | 100 | 100 | 83 | 100 |
| 16 | 54 | 18/18 | 18 | 18/18 | 34/36 | 88 | 100 | 83 | 100 | 100 |
| 17 | 108 | 17/18 | 18 | 18/18 | 33/36 | 73* | 100 | 100 | 100 | 100 |
| 18 | 54 | 18/18 | 18 | 18/18 | 28/36 | 75* | 67* | 80 | 83 | 83 |
| 19 | 18 | 17/18 | 18 | 8/18-18/18 | 36/36 | 100 | 100 | 100 | 100 | 100 |
| 20 | 36 | 18/18 | 18 | 18/18 | 33/36 | 92 | 100 | 100 | 67* | 100 |
| 21 | 36 | 18/18 | 18 | 18/18 | 36/36 | 100 | 100 | 100 | 100 | 100 |
| 22 | 36 | 18/18 | 18 | 18/18 | 36/36 | 100 | 100 | 100 | 100 | 100 |
| 23 | 54 | 18/18 | 18 | 18/18 | 36/36 | 100 | 100 | 100 | 100 | 100 |
| 24 | 18 | 18/18 | 18 | 18/18 | 34/36 | 83 | 100 | 100 | 100 | 100 |

The last four columns represent accuracy of responding (within 5 s) in the equivalence test across the five trial types: stimulus equivalence trials (SE) or one of the four distractor trial types—phonologically and orthographically similar (PO), phonologically similar only (P), orthographically similar only (O), and neither orthographically nor phonologically similar (N). Asterisks denote performances that failed to reach the mastery criterion

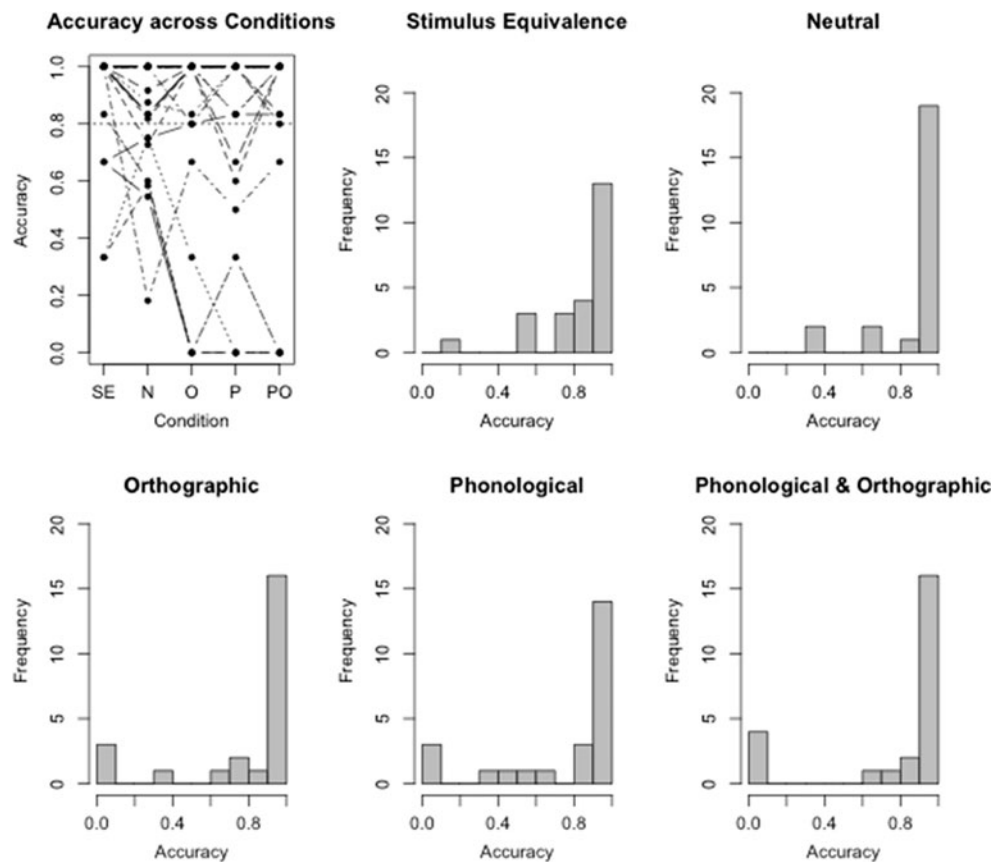
this did not seem to adversely affect their performance during A–C testing, with only 1 participant failing to demonstrate the required discriminations on the first block of testing (participant 19). This participant demonstrated the required performance on exposure to a second block of training and testing.

Participants were deemed to have failed to demonstrate equivalence-consistent responding in a trial type if they produced fewer than 80 % correct responses. During some trials, poor communication between the Wii remote and the computer resulted in data loss during trials, and these trials were excluded from further analyses (5.6 % of all trials). Eighty percent was employed as a threshold, since this allowed for no more than one incorrect response in the distractor conditions (six trials were presented in these conditions). Seven (29 %) participants produced fewer than 80 % correct responses on the stimulus equivalence trials, 5 (21 %) on the combined PO similarity trials, 7 (29 %) on the P trials, 5 (21 %) on the O trials, and 4 (17 %) on the N trials. Figure 4 presents the distribution of equivalence responses across conditions. In the top right panel of the figure, the plot shows that accuracy decreased as the incorrect comparison stimulus increased in visual and phonological similarity to the sample. This effect was most pronounced for those participants who did not demonstrate the equivalence performance in the stimulus equivalence

condition. The 17 participants who produced greater than 80 % correct responses on the stimulus equivalence trials produced, on average, 96 % correct responses on the stimulus equivalence trials, 100 % correct on the N trials, 98 % correct on O trials, 94 % correct on the P trials, and 98 % correct on the PO trials. In contrast, the 7 participants who produced fewer than 80 % correct responses on the stimulus equivalence trials produced, on average, 59 % correct responses on the stimulus equivalence trials, 69 % correct on the N trials, 40 % correct on the O trials, 38 % correct on the P trials, and 35 % correct on the PO trials. One explanation of this effect is that if responding in accordance with the equivalence relation was not strong (or present at all), the distractor stimuli were more likely to exert control over behavior.

The remaining five panels in Fig. 4 provide histograms of the average accuracy of participants in each condition. For the top center panel, it can be seen that, even though 7 participants failed to achieve the mastery criterion in the stimulus equivalence condition, accuracy was less variable in that condition than in the O, P, and PO conditions (bottom panels). From the histograms in the bottom panel of the figure, it can be seen that a number of participants solely chose the distractor stimuli, making these distributions more bimodal than those observed in the stimulus equivalence and N conditions. A generalized linear model based on the binomial distribution was employed to investigate whether

Fig. 4 Graphical representations of the distribution of accurate (equivalence) responses during phase 5 B–C testing. The top left panel shows the accuracy of participants across conditions. In the remaining five panels, histograms depict the frequency of accuracy scores within each condition. For example, in the stimulus equivalence condition, 13 participants made 90 %–100 % correct responses, 4 made 80 %–90 %, and so on



the distributions of accurate (1) and inaccurate (0) responses were significantly different across trial types. This model was run in the R statistical analysis environment (R Development Core Team, 2011) using the R packages lme4 (Bates, Maechler, & Bolker, 2011) and languageR (Baayen, 2011). Statistical analyses based on a normal distribution (such as an ANOVA) would not have been suitable for such contrasts (see Jaeger, 2008, for a detailed discussion). Participant (variance = 7.3316) and condition within participant (1.4347) were included in the model as random variables (Pinheiro & Bates, 2000, p. 23). Accuracy was compared across conditions using the N condition as reference. Accuracy was significantly lower in all conditions than in the neutral distractor condition: stimulus equivalence, $b = -1.501$, $z = -2.101$, $p = .0357$; orthographic, $b = -1.780$, $z = -2.310$, $p = .021$; phonological, $b = -2.336$, $z = -3.075$, $p = .002$; and phonological–orthographic $b = -1.944$, $z = -2.510$, $p = .012$. To test whether there were any differences among the interference conditions, the model was rerun with the phonological trial type as the reference (all interference conditions were significantly different from neutral, and phonological was most different), and no significant differences were obtained other than with the N condition.

At this point, to compare characteristics of correct responses across conditions, all 7 participants who produced fewer than 80 % equivalence responses in the stimulus equivalence condition were excluded. A number of further exclusion criteria were employed. All responses that took longer than 5 s were removed from the analysis. Response latencies longer than 5 s were assumed to be due to extraneous variables (e.g., adjusting to the response requirements of the final test phase, distraction, and so on). Following visual inspection, the trajectories of 1 participant (participant 20) exhibited unusual patterns (repeatedly flicking the Wii remote from side to side), so his data were removed from further analysis. Data from 500 trials across 16 participants were included in the following analyses.

Four dependent variables were employed to investigate potential effects of competing stimulus control (see Table 3). Reaction time was divided into latency and movement time. On each trial, a 100-pixel square escape region was created

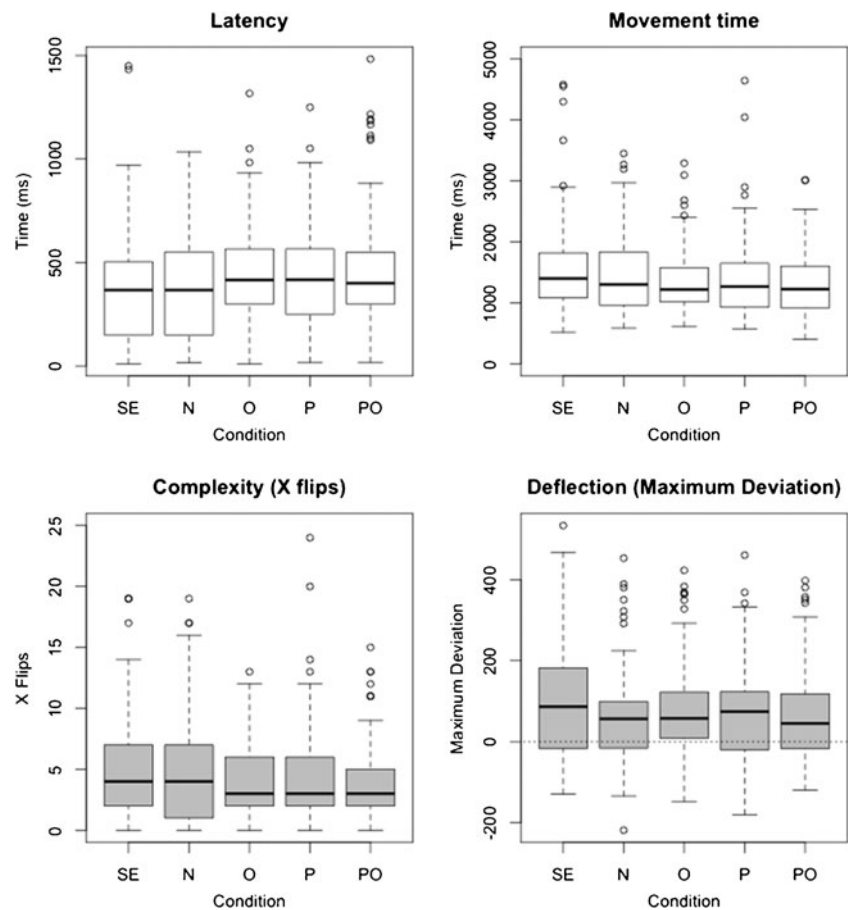
around the point at which the participant pressed to choose the sample stimulus (see Fig. 6). The time taken to leave this escape region provided the latency measure. Movement time referred to the time taken to move the cursor from the escape region to 100 ms prior to choosing a comparison stimulus. Complexity was measured in terms of x flips (Dale et al., 2008), the number of times during a response that a participant changed direction on the x -axis. Put simply, a movement to the left and then to the right constituted one x flip. Finally, the maximum deviation from the straight line from boundary of the escape region to the chosen comparison was employed to index the degree of deflection of the response trajectory from the shortest trajectory (Dale et al., 2008). Both x flip and maximum deviation calculations were performed on the trajectory outside of the 100-pixel square escape region. Graphical representations of the variability in the dependent variables across conditions are presented in Fig. 5.

There were no significant differences in the response latency, movement time, or complexity (x flips) of response trajectories during N trials and distractor trials. Linear mixed effects models (fit using restricted maximum likelihood) were employed to test for statistically significant differences in these dependent variables across conditions. These analyses are similar to repeated measures ANOVAs but are more powerful and more flexible when multiple values are available per condition (see Pinheiro & Bates, 2000). Response latency was relatively similar across conditions, although there seemed to be less variability in the distractor conditions than in either the stimulus equivalence trials or the N trials (see Fig. 5, upper left panel). Latency was log transformed to reduce skew. Participant (variance = 0.062) was included in the model as a random variable, but condition within participant was omitted, since it did not significantly improve the model. Latency was compared across conditions, using the neutral condition as reference. No significant differences were found between the N condition and the other conditions: orthographic, $b = 0.166$, $t = 0.91$, $p > .05$; phonological, $b = 0.115$, $t = 0.62$, $p > .05$; phonological–orthographic, $b = 0.265$, $t = 1.48$, $p > .05$; or stimulus equivalence, $b = -0.022$, $t = -0.14$, $p > .05$.

Table 3 Means and standard deviations (in parentheses) of the four measures of response trajectories employed in phase 5 B–C testing

| Condition | Latency (ms) | Movement time (ms) | Maximum deviation (pixels) | x flips (n) |
|---------------------------|--------------|--------------------|----------------------------|-------------------|
| Stimulus Equivalence | 375 (269) | 1,534 (704) | 106 (142) | 5.31 (4.2) |
| <i>Distractors</i> | | | | |
| Neutral | 390 (280) | 1,469 (669) | 64 (116) | 4.89 (4.4) |
| Orthographic | 412 (264) | 1,386 (574) | 80 (127) | 4.28 (3.2) |
| Phonological | 420 (280) | 1,414 (708) | 74 (119) | 4.63 (4.3) |
| Phonological–orthographic | 446 (310) | 1,338 (542) | 73 (119) | 4.18 (3.1) |

Fig. 5 Box plots of four dependent variables measured during phase 5 B–C testing. In all panels, higher (positive) scores denote greater competition



Movement time was also relatively similar across conditions (see Fig. 5, upper right panel). Following log transformation, movement time was compared across conditions using the neutral condition as reference. No significant differences were found between the neutral condition and the other conditions: orthographic, $b = -0.029$, $t = -0.48$, $p > .05$; phonological, $b = -0.039$, $t = 0.65$, $p > .05$; phonological–orthographic, $b = -0.069$, $t = -1.17$, $p > .05$; or stimulus equivalence, $b = 0.061$, $t = 1.17$, $p > .05$. As in the latency model, participant (variance = 0.020) was included in the model as a random variable, but condition within participant was excluded because it did not significantly improve the model.

There were no significant differences in the complexity of trajectories from the sample to the chosen comparison stimulus as measured in x flips between the N condition and the other conditions: orthographic, $b = -0.593$, $t = -1.021$, $p > .05$; phonological, $b = -0.296$, $t = -0.503$, $p > .05$; phonological–orthographic, $b = -0.711$, $t = -1.239$, $p > .05$; or stimulus equivalence, $b = 0.475$, $t = 0.935$, $p > .05$. Once more, participant was included as a random variable (variance = 0.984).

Average trajectories are provided in Fig. 6. All trajectories are shown as though choosing a stimulus in the top right-hand corner of the screen. In order to do this, trials on

which participants chose stimuli in the top left-hand corner were reflected in the y -axis at zero. Average trajectories were created by calculating the 20 % trimmed mean (the

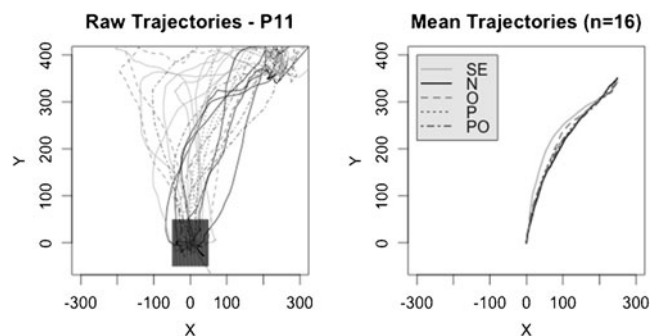


Fig. 6 Wii remote trajectories across conditions. All trajectories end on the right; leftward choices were reflected in the y -axis at 0,0 to facilitate comparison. The left panel depicts whole trajectories of 1 participant. Considerable jitter can be observed around the sample stimulus where the participant maintained the cursor position against gravity and the downward deflection induced by pressing the button on the Wii remote to select the sample stimulus prior to moving. A 100-pixel square area constituted an “escape region” (shaded section). The right panel shows average trajectories (trimmed mean; see the text for details) across conditions from the “escape region” to 100 ms prior to the final choice response. X flips and deviation scores were calculated on the basis of trajectories within these limits

lower and upper fifths of the distribution were excluded to rule out extreme scores) for x and y at each of the 101 time steps in each condition and then combining the resulting vectors. The stimulus equivalence trial trajectories exhibited the greatest deflection toward the incorrect stimulus. In order to statistically compare deflection of trajectories across conditions, the maximum deviation from the straight line between the start and end of the trajectory was calculated for each trajectory. As before, a linear mixed model analysis was employed with participant (variance = 0.920) as random variable and the N condition as reference. Maximum deviation was significantly greater in the stimulus equivalence condition than in the N condition, $b = 42.423$, $t = 2.717$, $p = .007$, but there were no significant differences between the N and the O, $b = 15.263$, $t = 0.857$, $p > .5$, P, $b = 7.274$, $t = 0.403$, $p > .5$, and PO, $b = 8.001$, $t = 0.454$, $p > .5$, conditions. To check whether there were any significant differences between the stimulus equivalence and distractor conditions, the analysis was rerun with the stimulus equivalence condition as reference. Maximum deviation was significantly greater in the stimulus equivalence condition than in the P, $b = -35.15$, $t = -2.190$, $p = .029$, and PO, $b = -34.42$, $t = -2.213$, $p = .027$, conditions. The difference between the stimulus equivalence condition and the O condition was not significant, $b = -27.16$, $t = -1.721$, $p = .086$. Probabilities were estimated using Markov chain Monte Carlo simulation (LanguageR; Baayen, 2011).

Discussion

The present study supplements a body of research (e.g., Stewart et al., 2002; Tyndall et al., 2004) that has demonstrated that topographical and functional similarity of incorrect comparisons may induce responding that competes with responding in accordance with equivalence relations. The probability of equivalence-consistent responding was significantly lower in the presence of phonologically similar, orthographically similar, and both phonologically and orthographically similar incorrect stimuli than in traditional stimulus equivalence trials or in the presence of neutral novel stimuli. However, the effect of distractor stimuli was most pronounced when individuals failed to demonstrate equivalence in the traditional equivalence trials. In fact, when participants produced over 80 % equivalence-consistent responding in traditional equivalence trials, the presence of phonological and orthographic distractor stimuli did not reduce accuracy.

In order to further investigate response competition, we analyzed the response trajectories of correct responses by participants who demonstrated equivalence. We employed four dependent variables to compare the trajectories of responses during the various trial types. For three of these

dependent variables, there were no differences across conditions relative to the neutral condition: latency, movement time and x flips (complexity). The fourth dependent variable was the maximum deviation from the straight line from start to the end of a response, which was employed as an index of curvature of the response toward the incorrect stimulus. Response trajectories were significantly more curved on traditional stimulus equivalence trials than on the N, P, and PO trials (but not the O condition). This finding suggests that for those that demonstrated equivalence, the incorrect stimulus presented during traditional equivalence trials induced greatest response competition. One reason for this effect may be that choosing the incorrect comparison stimulus employed in the traditional equivalence trials had previously been reinforced during some MTS training trials. In contrast, none of the novel distractor stimuli had been presented during training, so choosing these stimuli had not previously been reinforced.

One way to conceptualize the findings of the present study is to consider the computer screen as a space within which the participant moves from the sample stimulus at the bottom of the screen to the appropriate comparison stimulus at the top. As the participant engages in the choice response, response competition is expressed not only in the relative probability of choosing either stimulus, but also in the manner in which the choice is made. Following Spivey et al. (2005) and Killeen (1992), one might consider the alternative choice locations to constitute “attractors” in behavioral space. A simple definition of an attractor is a location in space to which trajectories are drawn, and the strength of these attractors determines the trajectories in space that are observed. In the present study, for those participants who demonstrated equivalence, it seems that the previous history of reinforcement for choosing the incorrect comparison during training increased the strength (or depth) of that attractor. As a consequence, trajectories were drawn further from the correct response during traditional equivalence trials before, eventually, the trajectory was captured by the correct stimulus location. For participants who did not demonstrate equivalence, participants tended to choose either stimulus during equivalence and N trials, suggesting similar attraction to both stimuli. However, in the presence of P and PO incorrect comparisons, participants tended to choose these incorrect comparisons much more readily. Thus, when there was minimal or weak attraction to the correct (equivalent) stimulus, the P and PO stimuli exerted strong control over responding, even though they exerted negligible effects on the trajectories of those participants who demonstrated equivalence. These qualitatively different performances suggest that for those that acquired equivalence, a bifurcation in response patterns occurred, from responding influenced by preexperimental contingencies (choosing similar sounding stimuli) to responding influenced by

experimental contingencies (equivalence vs. previous experimental reinforcement). In the dynamical parlance inspired by Killeen and by Marr (1992), the response topographies reflect a “phase transition” in the acquisition of equivalence relations. Further research may seek to identify the “tipping point” in the strength of the equivalence responses, at which control shifts from preexperimental to experimental contingencies.

In the empirical investigation of derived stimulus relations, a focus on response topography may provide a fruitful test-bed for current and future theories. In the applied literature, fluency of verbal responding has been correlated with numerous measures that seem to indicate greater response strength (e.g., retention, endurance, application; Johnson & Layng, 1996; Lindsley, 1992). Response latency of accurate responses is often used as an index of fluency, but fluency might also be measured in terms of reduced complexity and variability of verbal response topographies. Refinement of topography may provide an index of refinement of stimulus control. In addition, the development of fluency within an experimental session might be observed in reduced curvature across trials. In the present study, we had too few test probes to test for changes in trajectory indices within the experimental session. For example, it is possible that previously unseen distractor stimuli may exercise stronger competing stimulus control when first encountered, and this effect may decrease across trials. Indeed, the presence of a “known” distractor might eventually facilitate accurate responding, through S– control (thanks to an anonymous reviewer for this suggestion). Future research might compare action dynamics of responding earlier and later within experimental sessions to explore these possibilities.

The present findings contribute to an understanding of the various sources of stimulus control in tests for stimulus equivalence. Randell and Remington (1999) demonstrated that phonological similarity of responses (*boat*, *goat*) in the presence of equivalence class members (pictures of boats and goats) facilitated equivalence-consistent responding. When phonological similarities were not in line with equivalence relations, equivalence-consistent responding was less likely to be observed. In the present study, participants who demonstrated equivalence in the traditional equivalence trials reliably avoided the phonologically and orthographically similar incorrect comparisons. Of 290 trials, in only 10 (3.4 %) did the participant choose the incorrect orthographically or phonologically similar comparison stimulus. In contrast, participants who failed to demonstrate equivalence on traditional equivalence trials chose these incorrect stimuli (O, P, and PO distractors) on 71 of 118 trials (60.1 %). This contrast suggests that the strength of the rhyming effect may be context dependent. That is, when stimulus control has not previously been established, phonological and orthographic similarities may provide a basis for stimulus control, but

when stimulus control has already been established, stimulus control based on phonological and orthographic similarities between sample and comparison may be weak, if present at all. In Randell and Remington’s (1999) study, both sources of control contributed to the same outcome, whereas in the present study, these sources of control were in opposition. Further research is required to further investigate the conditions under which phonological responses emerge to facilitate or impede equivalence responding. For instance, researchers may seek to compare the effects of orthographic and/or phonological similarity on simple discriminations, by providing distractor stimuli similar to correct comparisons.

The present data, alongside those of Randell and Remington’s (1999) work, suggest that the probability or strength of private phonological responses varies across contexts. Indeed, there is a considerable literature on bilingualism (e.g., Spivey & Marian, 1999) that shows that bilinguals’ phonological responses in their different languages vary in strength and probability depending on linguistic context (e.g., for an English–French bilingual, *pain* rhymes with *wane* in an English context, but rhymes with *ban* in a French context). Spivey and Marian, for example, noted that Russian–English late bilinguals (participants who learned Russian as their first language for at least 16 years) provided with an English language instruction (e.g., “Pick up the marker”) were more likely to look at an incorrect picture stimulus (a stamp; *marku* in Russian) that was phonologically similar in Russian to the correct English choice (*marker–marku*), when the Russian language was more salient in the experimental context (e.g., they were greeted in Russian at the beginning of the experiment). Such context sensitivity is explicitly predicted by SCTC and RFT. RFT suggests that contextual variables influence the strength and probability of relations between stimuli (Crel) and the strength and probability of responses (Cfunc) occasioned by those stimuli (see Wulfert & Hayes, 1988, for an empirical demonstration). The RFT framework thus includes the context dependency of verbal responses, and once potential contextual variables are identified, specific predictions may be made (e.g., Steele & Hayes, 1991; Wulfert & Hayes, 1988). Future studies may seek to identify and test naturally occurring contextual variables or to train novel contextual variables that will act as Cfuncs and Crels in experimental environments. In line with this operant interpretation, recent work (Bartolotti & Marian, 2012) suggests that bilinguals’ experience of multiple languages reduces cross-language interference (expressed in computer mouse and eye movements) when learning new languages, relative to monolinguals.

A variety of conditional discrimination procedures have been employed to train and test stimulus equivalence (for reviews, see Arntzen & Holth, 1997; Saunders & Green,

1999). The yield (i.e., the number of participants who demonstrated equivalence; 70.8 % in the present study) using our training procedure was lower than it may have been had we repeatedly retrained participants, interspersed baseline trials with equivalence trials, or gradually faded reinforcement for baseline trials. In fact, in the latter case, Arntzen, Grondahl, and Eilifsen (2010) provided evidence that gradually fading feedback does not account for differential yields in equivalence training paradigms. Nevertheless, even though reduced yield would not account for the differences observed between trial types in stimulus equivalence testing in the present study, future studies might investigate whether procedures that produce greater yields also produce faster, more direct equivalence responses. In addition, the present procedures employed two comparison stimuli when testing for equivalence, and there is some controversy surrounding the use of such procedures. For example, Carrigan and Sidman (1992) advocated that researchers who employ two-comparison conditional discriminations should explicitly control whether the positive or negative comparison governs their participants' choices or use a minimum of three comparison stimuli. That is, it is possible, in two-choice procedures, for participants to choose the appropriate stimulus by avoiding the incorrect stimulus (i.e., through exclusion). However, Boelens (2002) provided a robust rebuttal of these arguments. In addition, more recently, Minster et al. (2011), using a three-choice design, provided class-unique incorrect comparison stimuli during testing and found no evidence of control by exclusion in tests for equivalence (instead of reliably avoiding these stimuli, the majority of participants chose them). Nevertheless, analyses of response trajectories such as those in the present study might provide novel approaches to contrasting S+ and S− control in future research.

In the context of the present study, a two-choice procedure was employed to encourage attending to the incorrect comparison in testing. In traditional tests of stimulus equivalence, it is the relationship between the sample and comparison that is of interest. Thus, it may not be crucial that the incorrect comparison is discerned. In contrast, in the present study, it was of particular importance that the incorrect comparison was attended to, in order to identify any conflicting stimulus control due to the functional or topographical similarity of the incorrect comparison with the sample. If a three- or four-choice procedure were used, it would be more likely that the participant would not attend to any one of the two or three incorrect comparisons. If non-similar comparison stimuli were included in these trials, it would be more likely that the participants might fail to attend to the critical comparison stimulus. Furthermore, the combinations of phonologically and orthographically similar words employed in the present study are relatively rare, so it would be difficult to find multiple phonologically and

orthographically similar incorrect comparisons for each of the trial types in the present study.

From Sidman's earliest studies of stimulus equivalence, the most evident application has been in the establishment and enhancement of reading behavior. Equivalence-consistent responses exhibit very similar characteristics to semantic or symbolic relations between stimuli (Barnes, 1994; Dougher, 1998; Sidman, 1994, 2000). For this reason, features of stimulus objects or relevant behavioral histories that serve to occasion responses that reduce the probability of equivalence-consistent responding may also serve to impair reading performance. As the present study indicates, orthographic and phonological similarity constitute two sources of such potential competing stimulus control. This is particularly important for children learning to read the English language, in which the low level of orthographic consistency (how stable the relationship is between the textual stimulus and the appropriate phonological stimulus/response) leads to higher rates of mispronunciations in young English speakers than in speakers of other, more regular languages (Frith, Wimmer, & Landerl, 1998). For young readers, developing tightly discriminated phonological responses facilitates word identification and the probability of equivalence-consistent responses by minimizing potential conflicts in stimulus control. There is little doubt that effective reading instruction programs already focus on the development of tightly controlled rapid discriminations of textual stimuli. Future studies may seek to further the investigation of reading behavior by developing laboratory analogs of such effective interventions.

Author's Note The authors wish to thank Randolph Grace and two anonymous reviewers for their insightful comments on the manuscript. This research was supported by the award of a National University of Ireland Galway Millennium Grant to the first author. A portion of these data was submitted as part of the undergraduate thesis of the third author. Address correspondence to Denis O'Hora, School of Psychology, National University of Ireland, Galway, Ireland (email: denis.ohora@nuigalway.ie).

References

- Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *Psychological Record*, *54*(2), 275–291.
- Arntzen, E., & Holth, P. (1997). Probability of stimulus equivalence as a function of Training Design. *Psychological Record*, *47*(2), 309–320.
- Arntzen, E., Grondahl, T., & Eilifsen, C. (2010). The effects of different training structures in the establishment of conditional discriminations and subsequent performance on tests for stimulus equivalence. *Psychological Record*, *60*(3), 437–462.
- Audacity Development Team (n.d.). Audacity [software]. Available from <http://audacity.sourceforge.net/>
- Baayen, R. H. (2011). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics"

- (R package version 1.2) [software]. Available from <http://CRAN.R-project.org/package=languageR>
- Barnes, D. (1994). Stimulus equivalence and relational frame theory. *Psychological Record*, 44(1), 91–124.
- Bartolotti, J., & Marian, V. (2012). Language learning and control in bilinguals and monolinguals. *Cognitive Science*, 36(6), 1129–1147.
- Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes (R package version 0.999375-40) [software]. Available from <http://CRAN.R-project.org/package=lme4>
- Boelens, H. (2002). Studying stimulus equivalence: Defense of the two-choice procedure. *The Psychological Record*, 52(3), 305–314.
- Bonatti, L. L. (n.d.). Psyscope X (Build 51) [software]. Available from <http://psy.cns.sissa.it>
- Carrigan, P. F., & Sidman, M. (1992). Conditional discrimination and equivalence relations: A theoretical analysis of control by negative stimuli. *Journal of the Experimental Analysis of Behavior*, 58(1), 183–204.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, 25(2), 257–271.
- Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed processing approaches. *Psychological Review*, 100, 589–608.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). The DRC model: A model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–258.
- Dale, R., Roche, J., Snyder, K., & McCall, R. (2008). Exploring action dynamics as an index of paired-associate learning. *PLoS One*, 3, e1728.
- R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria [software]. Available from <http://www.R-project.org/>
- Dougher, M. J. (1998). Stimulus equivalence and the untrained acquisition of stimulus functions. *Behavior Therapy*, 29(4), 577–591.
- Dube, W. V., & McIlvane, W. J. (2003). Stimulus Control Topography Coherence Theory: Foundations and extensions. *Behavior Analyst*, 26(2), 195–213.
- Dugdale, N., & Lowe, C. F. (1990). Naming and stimulus equivalence. In D. E. Blackman & H. Lejeune (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 115–138). Hove, England: Erlbaum.
- Frith, U., Wimmer, H., & Landerl, K. (1998). Differences in phonological recoding in German- and English-speaking children. *Scientific Studies of Reading*, 2(1), 31–54.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition*. New York: Plenum Press.
- Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior*, 65(1), 185–241.
- Horne, P. J., & Lowe, C. F. (1997). Toward a theory of verbal behavior. *Journal of the Experimental Analysis of Behavior*, 68(2), 271–296.
- Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59(4), 434–446.
- Johnson, K. R., & Layng, T. V. J. (1996). On terms and procedures: Fluency. *Behavior Analyst*, 19(2), 281–288.
- Killeen, P. R. (1992). Mechanics of the animate. *Journal of the Experimental Analysis of Behavior*, 57(3), 429–463.
- Kimura, H. (n.d.). Darwiin Remote [software]. Available from <http://sourceforge.net/projects/darwiin-remote/>
- Leslie, J. C., Tierney, K. J., Robinson, C. P., Keenan, M., Watt, A., & Barnes, D. (1993). Differences between clinically anxious and non-anxious subjects in a stimulus equivalence training task involving threat words. *Psychological Record*, 43(1), 153–161.
- Lindsley, O. R. (1992). Precision teaching: Discoveries and effects. *Journal of Applied Behavior Analysis*, 25(1), 51–57.
- Lowe, C. F., Horne, P. J., Harris, F. D. A., & Randle, V. R. L. (2002). Naming and categorization in young children: Vocal tact training. *Journal of the Experimental Analysis of Behavior*, 78(3), 527–549.
- Luciano, C., Becerra, I. G., & Valverde, M. R. (2007). The role of multiple-exemplar training and naming in establishing derived equivalence in an infant. *Journal of the Experimental Analysis of Behavior*, 87(3), 349–365.
- Mandell, C., & Sheen, V. (1994). Equivalence class formation as a function of the pronounceability of the sample stimulus. *Behavioural Processes*, 32(1), 29–46.
- Marr, M. J. (1992). Behavior dynamics: One perspective. *Journal of the Experimental Analysis of Behavior*, 57(3), 249–266.
- Miguel, C. F., Petursdottir, A. I., Carr, J. E., & Michael, J. (2008). The role of naming in stimulus categorization by preschool children. *Journal of the Experimental Analysis of Behavior*, 89(3), 383–405.
- Minster, S. T., Elliffe, D., & Muthukumaraswamy, S. D. (2011). Emergent stimulus relations depend on stimulus correlation and not on reinforcement contingencies. *Journal of the Experimental Analysis of Behavior*, 95(3), 327–342.
- Minster, S. T., Jones, M., Elliffe, D., & Muthukumaraswamy, S. D. (2006). Stimulus equivalence: Testing Sidman's (2000) theory. *Journal of the Experimental Analysis of Behavior*, 85(3), 371–391.
- Moxon, P. D., Keenan, M., & Hine, L. (1993). Gender-role stereotyping and stimulus equivalence. *Psychological Record*, 43(3), 381–393.
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer.
- Plaud, J. J. (1995). The formation of stimulus equivalences: Fear-relevant versus fear-irrelevant stimulus classes. *Psychological Record*, 45(2), 207–222.
- Plaud, J. J. (1997). Behavioral analysis of fear-related responding using a modified matching-to-sample procedure. *Scandinavian Journal of Behavior Therapy*, 26(4), 157–170.
- Plaut, C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56–115.
- Polich, J., McCarthy, G., Wang, W. S., & Donchin, E. (1983). When words collide: Orthographic and phonological interference during word processing. *Biological Psychology*, 16(3–4), 155–180.
- Randell, T., & Remington, B. (1999). Equivalence relations between visual stimuli: The functional role of naming. *Journal of the Experimental Analysis of Behavior*, 71(3), 395–415.
- Randell, T., & Remington, B. (2006). Equivalence relations, contextual control, and naming. *Journal of the Experimental Analysis of Behavior*, 86(3), 337–354.
- Saunders, R. R., & Green, G. (1999). A discrimination analysis of training-structure effects on stimulus equivalence outcomes. *Journal of the Experimental Analysis of Behavior*, 72(1), 117.
- Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, 14(1), 5–13.
- Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative, Inc.
- Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, 74(1), 127–146.
- Sidman, M., & Cresson, O. (1973). Reading and the cross-modal transfer of stimulus equivalences in severe retardation. *American Journal of Mental Deficiency*, 77(5), 515–523.

- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching-to-sample: An extension of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37(1), 5–22.
- Skinner, B. F. (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 10393–10398.
- Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, 10(3), 281–284.
- Steele, D., & Hayes, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior*, 56(3), 519–555.
- Stewart, I., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2002). Stimulus equivalence and nonarbitrary relations. *Psychological Record*, 52(1), 77–88.
- Tonneau, F. (2001). Equivalence relations: A critical analysis. *European Journal of Behavior Analysis*, 2(1), 1–33.
- Tonneau, F., & Gonzalez, C. (2004). Function transfer in human operant experiments: The role of stimulus pairings. *Journal of the Experimental Analysis of Behavior*, 81(3), 239–255.
- Tyndall, I. T., Roche, B., & James, J. E. (2004). The relation between stimulus function and equivalence class formation. *Journal of the Experimental Analysis of Behavior*, 81(3), 257–266.
- Watt, A., Keenan, M., Barnes, D., & Cairnes, E. (1991). Social categorization and stimulus equivalence. *Psychological Record*, 41(1), 33–50.
- Wilson, M. D. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, 20(1), 6–11.
- Wulfert, E., & Hayes, S. C. (1988). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior*, 50(2), 125–144.